

## **Data Management Plan (DMP) Title:** Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative

**Last update:** 24 February 2025

Note: This DMP meets new formatting requirements of NOAA GOMO, replacing the previous NOAA DMP Template v2.0.1 used for version 05.

Does this DMP sit under a higher level DMP (\*Master DMP)? NO

Name of higher-level DMP(s) (if applicable): N/A

**DMP Type:** Discrete dataset(s)

### **1. General Description of Data to be Managed:**

#### **1.1. Name of the Data, data collection Project, or data-producing Program (Title):**

Shipboard Automated Meteorological and Oceanographic System (SAMOS) Initiative.

#### **1.2. Project purpose and summary description of the data:**

Since 2005, the SAMOS initiative has collected 1-minute average underway navigational, meteorological, and oceanographic observations (derived from higher frequency, several per minute up to 1 Hz, sensor measurements) from 45 research vessels. As of 2025, 30 vessels are actively contributing to the Initiative. Vessel selection is driven primarily by funding source (e.g., NOAA, NSF, and the Schmidt Ocean Institute); however, when possible we prioritize vessel that are (a) global operators as opposed to coastal vessels, (b) operated in remote ocean regions (e.g., Southern and Arctic oceans), and (c) routinely make simultaneous meteorological observations required to estimate air-sea fluxes. One-minute average observations are produced at 1-minute intervals onboard each participating vessel and are delivered in daily ship-to-shore email messages to the Marine Data Center (MDC) at the Florida State University (FSU). A SAMOS consists of a computerized data logging system that continuously records navigation (ship's position, course, speed, and heading); meteorological (winds, air temperature, pressure, moisture, rainfall, and radiation); and near ocean surface (sea temperature and salinity) parameters while a vessel is underway (A full list of observations requested from recruited vessels is available at <https://samos.coaps.fsu.edu/html/parameters.php>).

SAMOS data are collected, quality-evaluated, distributed, and archived to support a wide range of secondary data user communities (i.e., users not sailing on a given cruise). The data are essential to meet objectives for satellite algorithm development, calibration, and evaluation; numerical model evaluation and development; creating accurate estimates of air-sea fluxes; atmospheric and oceanic process studies; and a range of geoinformatics research and development.

1.2.1. List the name of each dataset included in this DMP (higher level and sub-level DMPs do not need to be included) or the name of the collection of data sets:

Dataset Name
SAMOS

1.2.2. DOI (if applicable) for each dataset:

Dataset Name	DOI(s)
SAMOS	<a href="https://doi.org/10.7289/v5qj7f8r">https://doi.org/10.7289/v5qj7f8r</a>

1.3. **Is this a one-time data collection, or an ongoing series of measurements:**

SAMOS is an ongoing series of measurements, updated daily via satellite communication/email service between the vessel operator and the MDC at FSU.

1.3.1. What is the data submission timeframe for ongoing series (Annually, Monthly, Weekly, Twice a Week, etc.)?:

Dataset Name	Data Submission Timeframe
SAMOS	Monthly (submitted to NCEI)

1.4. **Actual or planned temporal coverage of the data:**

Dataset Name	Duration	Frequency
SAMOS	May 2005 to present	1-min (averaged from higher frequency measurements over duration of each SAMOS-equipped ship expedition plus, in some cases, data collected when the ship is in port)

1.5. **Actual or planned geographic coverage of the data:**

Global Oceans (-90 to 90 N, 0 to 360 E), with highest sampling density in the oceans surrounding North America.

1.6. **Is there a data acquisition plan:** (yes/no) NO

1.6.1. Does acquisition include restricted designations: N/A

1.6.2. If yes, please explain: N/A

1.7. **Type(s) of data:**

(e.g., digital numeric data, imagery, photographs, video, audio, database, tabular data, etc.)

Dataset Name	Data Type	Format (netcdf, ascii, etc.)
SAMOS	Digital numeric data (and associated metadata)	netCDF

1.8 **Approximate data volume:**

Dataset Name	Data Volume Estimate
SAMOS	current database: ~22 Gb average daily increase: ~10 Mb

**1.9. Data collection method(s):**

(e.g., satellite, airplane, unmanned aerial system, radar, weather station, moored buoy, research vessel, autonomous underwater vehicle, animal tagging, manual surveys, enforcement activities, numerical model, etc.)

Dataset Name	Assets required	Method by which data are collected or generated
SAMOS	Research Vessels	Underway in situ sensors

1.10. If data are from a NOAA Observing System of Record, indicate name of system:  
Voluntary Observing Ship

1.10.1. If data are from another observing system, please specify:  
Ships of Opportunity Programme, GO-SHIP, Ship Observation Team

**2. Point of Contact for this Data Management Plan (author or maintainer)**

- 2.1. Name: Mr. Shawn R. Smith
- 2.2. Title: Senior Research Associate
- 2.3. Affiliation or facility: MDC, Center for Ocean-Atmospheric Prediction Studies (COAPS), FSU
- 2.4. E-mail address: srsmith@fsu.edu
- 2.5. Phone number: 850-644-6918

**3. Responsible Party for Data Management:**

Program Managers, or their designee, shall be responsible for ensuring the proper management of the data produced by their Program. Please indicate the responsible party below.

- 3.1. Position Title: Senior Research Associate
- 3.2. Name of current position holder: Mr. Shawn R. Smith
- 3.3. Email of current position holder: srsmith@fsu.edu

**4. Resources:**

Programs must identify resources within their own budget for managing the data they produce.

4.1. Have resources for management of these data been identified?:

Yes. Base funding for the SAMOS data center is provided by the National Oceanic and Atmospheric Administration's (NOAA) Global Ocean Monitoring and Observing Program (FundRef number 100000192) and the Office of Marine and Aviation Operations via

subawards administered by the Northern Gulf of Mexico Cooperative Institute at the Mississippi State University. Support for Academic Research Fleet vessels' participation within SAMOS is provided by the National Science Foundation (NSF), Oceanographic Instrumentation and Technical Services Program via a subcontracts from the Lamont Doherty Earth Observatory (pre 2024) and the Scripps Institution of Oceanography (2025 – today). The Schmidt Ocean Institute (SOI) provides contract funding to include the *RV Falkor* and *Falkor (too)* in the SAMOS initiative.

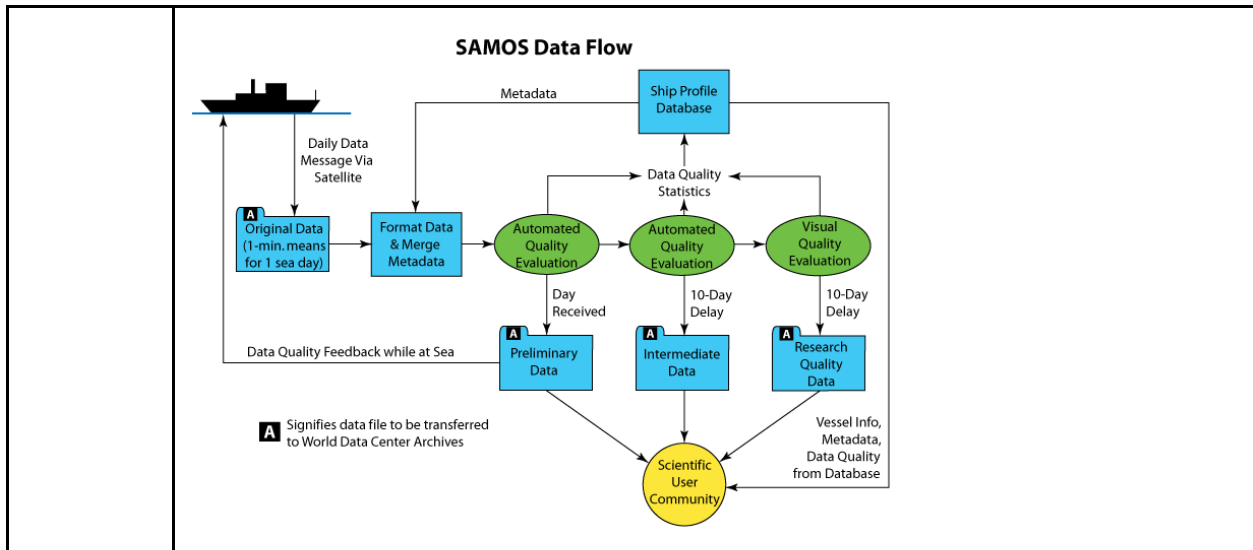
4.2. What is the approximate percentage of the overall project budget for these data devoted to data management to ensure compliance with data management requirements? (specify percentage or "unknown"): 90%

**5. Data Lineage and Quality:**

NOAA has issued Information Quality Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity of information which it disseminates.

5.1. Processing workflow of the data from collection or acquisition to making it publicly Accessible:

Dataset Name	Processing Workflow
SAMOS	<p>The flow of SAMOS observations from the vessel to the SAMOS data center begins with the operator sending all one-minute data records from the previous day to the MDC at 0000 UTC via an e-mail protocol (note: the vessels contributing to SAMOS from New Zealand and Australia post their data to a THREDDS server at the Australian Bureau of Meteorology and the MDC pulls the data from their server). SAMOS uses a custom key:value paired comma-separated value format for data transmission. Each operator encodes one-minute average observations, derived from higher sampling frequency instrumental observations, into the SAMOS format using their vessel's data acquisition software. Once received by the MDC, these observations are converted into a standard network common data form (netCDF) that is augmented with detailed ship and instrumental metadata provided to the MDC by each operator.</p> <p>At this point, the data undergo a series of scientific data quality control (QC) processes. The first QC process (see below) is fully automated and results in what the MDC calls a preliminary data file. On a 10-day delay from the observation date, intermediate files are automatically created by merging all preliminary files received for a given ship and observation day. This delay allows for receipt of delayed or corrected files from the RV. Finally, a select set of ships (including all recruited NOAA vessels) undergo visual QC to create a research-quality dataset. Preliminary, intermediate, and research quality netCDF files are made publicly accessible via the MDC as soon as they are produced via web, ftp, and THREDDS services (see below). Each month, the original data received from the vessel and all three levels of SAMOS-quality processed files are packaged for each ship and submitted to the National Centers for Environmental Information – Maryland (See figure below).</p>



5.1.1. If data at different stages of the workflow, or products derived from these data, are subject to a separate data management plan, provide reference to other plan and provide data source details: N/A

5.2. Quality control procedures employed:

Dataset Name	Measured Parameters (EOV/ECV)	QA/QC procedures and approving/accepting entity
SAMOS	<p><a href="#">See All Parameters</a> (includes non-EOV/ECV parameters)</p> <p>Primary (always measured):</p> <ul style="list-style-type: none"> <li>• Surface Wind Speed and Direction</li> <li>• Surface Atmospheric Pressure</li> <li>• Surface Air Temperature</li> <li>• Surface Water Vapor</li> <li>• Sea surface temperature</li> <li>• Sea surface salinity</li> </ul> <p>Secondary (desired, by not always measured):</p> <ul style="list-style-type: none"> <li>• Precipitation</li> <li>• Radiative fluxes (shortwave, longwave, photosynthetically available radiation)</li> <li>• fluorescence</li> <li>• oxygen</li> <li>• transmittance</li> </ul>	<p>The SAMOS QC system is based on the procedures used during the World Ocean Circulation Experiment (<a href="#">see original QC documentation</a>).</p> <p>The most recent documentation of the SAMOS QC procedures is available from Smith, S. R., K. Briggs, M. A. Bourassa, J. Elya, and C. R. Paver, 2018: Shipboard automated meteorological and oceanographic system data archive: 2005–2017. <i>Geosci Data J.</i>, 5, 73–86. <a href="https://doi.org/10.1002/gdj3.59">https://doi.org/10.1002/gdj3.59</a></p>

## **6. Data Documentation (Metadata):**

The Data Documentation Procedure requires that NOAA data be well documented, specifies the use of ISO 19115-2: 2009 and related standards for documentation of new data, and provides links to resources and tools for metadata creation and validation.

### 6.1. Process for producing and maintaining metadata, including organizations or facilities providing metadata hosting (describe or provide URL of description):

Dataset Name	Process for producing Metadata	Hosting Organization(s)
SAMOS	<p>SAMOS collects a broad suite of vessel- and instrument-specific metadata. The metadata specification was based on a combination of the metadata requirements for the Voluntary Observing Ship Climate (VOSCLim; <a href="http://hdl.handle.net/11329/7910.25607/OBP-1491">http://hdl.handle.net/11329/7910.25607/OBP-1491</a>) schema and metadata included in the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Freeman et al 2016). Metadata are submitted to the MDC from the operators initially when a vessel is recruited using forms available on the SAMOS website (<a href="https://samos.coaps.fsu.edu/html/participate.php">https://samos.coaps.fsu.edu/html/participate.php</a>). The operators update their metadata periodically by either resubmitting these forms to the MDC, providing updates during annual vessel device reviews, or via email when metadata changes occur. Metadata submitted to SAMOS are stored in a MySQL database, where all instrumental metadata are date tracked and versioned. This database is used to augment the SAMOS netCDF files that contain the actual navigational, meteorological, and oceanographic data.</p> <p>Once submitted to NCEI for archival, additional metadata records are prepared by NCEI for each archive package according to NCEI procedures and outlined in a data submission agreement between the MDC and NCEI. This includes creating FGDC and ISO-19115-2 metadata for each ship and month of SAMOS data submitted to NCEI.</p>	MDC, COAPS (original metadata)  NCEI (metadata archival)

### 6.2. Do you require assistance for producing and maintaining metadata? NO

## **7. Data Access:**

NAO 212-15 states that access to environmental data may only be restricted when distribution is explicitly limited by law, regulation, policy (such as those applicable to personally identifiable information or protected critical infrastructure information or proprietary trade information) or by security requirements. This restriction would include PII and other sensitive data (export

controlled data) and data restricted by contract or other written, binding agreement (permitted to be withheld under the Evidence Act) including commercial data licensed via contract, data obtained from another third party subject to a restrictive license (international partner, CRADA, etc.). The Data Access section of this Handbook contains specific guidance, recommends the use of open-standard, interoperable, non-proprietary web services, defines timeliness, provides information about resources and tools to enable data access.

7.1. Do these data comply with the general data access requirements?

Yes. The MDC at FSU has a long-standing policy to ensure free and open access to underway marine meteorological observations. For SAMOS, the policy ensures that all data provided to FSU will be redistributed to the user community and national archive centers without any restrictions or "holds." Data providers are requested to notify the SAMOS data center in writing (samoss@coaps.fsu.edu) if they have any concerns with this free and open data policy.

7.2. Intended data access method(s):

(Specify Web Service; API; FTP Bulk Download; HTTP Bulk Download; Website, web page, or portal; Asynchronous Ordering Service; To Be Determined; Unable to Provide Access; Limited Access Only; or No Access Intended)

Dataset Name	Intended Data Access Method(s)	Are data available through ERDDAP? (yes/no)
SAMOS	THREDDS; HTTP Bulk Download; FTP Bulk Download (All on <a href="#">SAMOS website</a> )	NO

7.2.1. Is the data, in part or in whole, restricted to the public?: (yes/no/TBD) NO

7.2.2. If Restricted, To Be Determined, or Unable to Provide Access, please explain: N/A

7.3. Name of NOAA organization or facility providing data access: N/A (Archival at NCEI)

7.3.1. If data hosting service (outside NOAA) is needed, please indicate URL of data access service or other methods, if known: [COAPS MDC](#)

7.3.2. URL(s) of data access service, if known:

Dataset Name	Data Access Links
SAMOS	<ul style="list-style-type: none"> <li>• THREDDS – <a href="https://www.coaps.fsu.edu/thredds-listing">https://www.coaps.fsu.edu/thredds-listing</a></li> <li>• HTTP –               <ul style="list-style-type: none"> <li>○ (By date): <a href="https://samoss.coaps.fsu.edu/html/data_availability.php">https://samoss.coaps.fsu.edu/html/data_availability.php</a>;</li> <li>○ (By cruise, when available): <a href="https://samoss.coaps.fsu.edu/html/cruise_data_availability.php">https://samoss.coaps.fsu.edu/html/cruise_data_availability.php</a></li> </ul> </li> <li>• FTP - <a href="ftp://ftp.coaps.fsu.edu/samos_pub/data">ftp://ftp.coaps.fsu.edu/samos_pub/data</a></li> </ul>

7.4. Tentative dissemination date by which data will be made publicly available per timeliness Requirement:

(Specify [Exact Date]; No Delay; one year from Collection; one year from QA/QC; or 2 years from QA/QC; Exceeds 2 years from QA/QC; To Be Determined; Unable to Provide Access; Limited Access Only; or No Access Intended)

Dataset Name	Tentative Dissemination Date or Delay until Data are Accessible
SAMOS	<ul style="list-style-type: none"> <li>• Preliminary Data = 1-day delay from measurement</li> <li>• Intermediate Data = 10-day delay from measurement</li> <li>• Research Quality Data = 10-day delay from measurement</li> </ul>

## **8. Data Preservation and Protection:**

The NOAA Procedure for Scientific Records Appraisal and Archive Approval<sup>14</sup> describes how to identify, appraise and decide what scientific records are to be preserved in a NOAA archive.

8.1. What is the actual or planned long term data archive location:

(Specify NCEI facility, or other specific archive, To Be Determined, Unable to Archive, or No Archiving Intended)

Dataset Name	Archive Location(s)
SAMOS	NCEI

8.1.1. If data preservation is to be provided by a data repository other than NOAA NCEI, please explain (e.g., data out-of-scope for NCEI, appraisal failed, preferred community of practice, cost considerations, technical considerations, etc.): N/A

8.1.2. If To Be Determined, Unable to Archive or No Archiving Intended, please clarify that selection: N/A

8.2. Data storage facility prior to being sent to an archive facility (if any):

MDC, COAPS, FSU

8.3. Approximate delay between data collection and submission to an archive facility:

Dataset Name	Delay until archival
SAMOS	6 weeks

8.4. How will the data be protected from accidental or malicious modification or deletion prior to receipt by the archive? Discuss data back-up, disaster recovery/contingency planning, and off-site data storage relevant to the data collection:

To protect original SAMOS data files and quality-controlled netCDF data from accidental or malicious modification or deletion before archival, we have implemented robust data



protection measures. These include daily automated backups, 4-hourly SQL data preservation, and ZFS snapshots for immutable backups. Versioning mechanisms, such as BitBucket, help track changes and restore previous versions. A disaster recovery plan is in place with periodic testing to ensure swift restoration, and backup copies are stored at separate campus locations to mitigate localized failures. Together, these measures ensure the security, integrity, and availability of the data until successful archival.

The archive packages submitted to NCEI also provide an off-site back-up of mission critical files. These packages include the original data received from the vessel and all versions of the quality-controlled data. Every file in the archive package is listed in a file manifest with MD5 checksums, which ensures that the data package is received at NCEI (and or returned to the MDC) without loss of any digital information.

### **9. Additional Line Office or Staff Office Questions:**

Line and Staff Offices may extend this template by inserting additional questions in this section.

9.1. If your project supports a cross-agency or international data collection effort, please include URL links to all NOAA-external plans, guidelines, or requirements that your data, metadata, or data management practices should/must follow or meet: N/A

#### 9.2. Data Usage

9.2.1. Identify all known products, web tools (visualization, decision support, etc.), assessments, or services that your collected data inform/support:

Dataset Name	Supported Products / Assessments
SAMOS	<ul style="list-style-type: none"><li>• International Comprehensive Ocean Atmosphere DataSet</li><li>• NOAA Surface Marine Underway Database</li><li>• MarineFlux</li><li>• NASA Cloud-based Data Match-up Service</li></ul>

9.2.2. Identify any additional users or anticipated users of your data:

Known users of SAMOS data are represented by the papers in the published literature.

See the following: <https://samos.coaps.fsu.edu/html/publications.php>

9.2.3. If data usage tracking is being done, provide relevant usage metrics implemented / to be implemented: Presently only tracking publications on a yearly basis that use/cite either the SAMOS Initiative or the SAMOS dataset.

9.3. Provide a brief description of any custom software or code developed or used for your project, and provide the source/link if available.

Software for this project includes a variety of custom codes developed in a research environment over the last 30 years. The code base includes:

- Automated email handling
- Automated metadata ingestion (for NOAA SCS5)
- Automated preprocessing of data received from vessel, metadata merge from database, and creation of initial netCDF files
- Automated quality control (creates preliminary product)
- Automated file merge with duplicate elimination (creates intermediate product)
- GUI for visual quality control (used to create research product)
- Scripts for archive packaging for pull by NCEI
- Numerous cron jobs

And the code base includes FORTRAN77, Perl, Python, IDL, C, PHP, etc. The code base is not standardized and not designed or documented for open-source distribution.